Claims

1. A method of identifying biologically active molecules from a set (S) comprising a predetermined number (N) of different molecules (M1, M2, ..., MN), said molecules being expected to be biologically active with respect to a predetermined target (T), each said molecule (M1, M2, ..., MN) of said set (S) being identified by a machine-readable descriptor (X1, X2, ..., XN), respectively, each said descriptor (X1,..., XN) being a vector with n vector elements (x1,..., xn), n being a natural number, each vector element (x1,..., xn) representing a predetermined molecular property, said method comprising the following steps:

a) selecting arbitrarily from said set (S) of molecules a subset (Su) comprising a predetermined first number (Nu) of molecules (Mi, ..., Mk);

c) assigning a fitness (fi, ..., fk), to each molecule (Mi, ..., Mk) of said subset (Su), respectively, said fitness (fi, ..., fk) being calculated according to a predetermined fitness measure f(X), said fitness measure f(X) being representative of the affinity of a molecule (Mi, ..., Mk) to said target (T);

m) establishing, according to a predetermined selection criterion (SC), from said subset (Su) a predetermined number (nc) of couples of molecules (MX, MY);

n) with each established couple of molecules: producing a predetermined number of descendant molecules (MO1, MO2) by recombining the

descriptors (X, Y) of said couple of molecules (MX, MY) according to a predetermined recombination scheme;

o) mutating each said descendant molecule (XM) by modifying the respective descriptor (XO) according to a predetermined mutation scheme (MS);

p) assigning a fitness (f) to each modified descendant molecule (MO), said fitness f(MO) being calculated according to the fitness measure f(X) of step b);

q) adding said modified molecules (MO) to said subset (Su);

r) removing a predetermined number of molecules from said subset (Su), the molecules to be removed being determined by a predetermined removal criterion (RC);

s) repeating steps b) to j) until a predetermined stop criterion (STC) is reached; and

t) outputting the subset (Su) of molecules according to step k).

2. The method according to claim 1, wherein said recombination scheme comprises weighted vector additions of the descriptors of each couple of molecules (MX, MY), whereby the sum of the respective weights is equal to unity.

3. The method according to claim 2, wherein said predetermined number of descendant molecules of step e) is two, and the weights for said vectorial

additions are p and 1 – p for producing the first descendant, and 1 – p and p for producing the second descendant, whereby $0 \leq p \leq 1$.

4. The method according to claim 1, wherein each descendant molecule (MO) which is not contained in said set (S) of molecules is replaced by the one molecule of said set having the smallest distance to said descendant molecule, said distance being calculated according to a predetermined metric criterion (MC).

5. The method according to claim 1, wherein said recombination scheme comprises combining a predetermined number of vector elements from the first descriptor (X) with a predetermined number of vector elements from the second descriptor (Y).

6. The method according to claim 1, wherein, if a modified descendant does not correspond to a molecule comprised in the set (S) of molecules, the fitness of said descendant molecule is calculated by using the descriptor (X) of the molecule of said set (S) having the smallest distance to said modified descendant descriptor, to a predetermined descriptor according to a predetermined metric criterion (MC).

7. The method according to claim 1, wherein said metric criterion MC is defined by:

$$\mathrm{MC}(X,Y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

with

x_i: vector element of said first descriptor X,

y_i: vector element of said second descriptor Y,

n: number of vector elements of said first and second descriptor, respectively.

8. The method according to claim 1, wherein said selection criterion (SC) is of the Roulette Wheel type wherein the probability (q) of selection of a molecule (M) is related to its fitness f(M).

9. The method according to claim 1, wherein said fitness values f of said descriptors (X) are scaled by

$$scal(f(X)) = a \cdot f(X) + b$$

with a, b being constants.

10. The method according to claim 1, wherein said mutation scheme is defined by addition of a random value r_Φ to each said vector element (x_i), said random value characterized by a probability density distribution (Φ) with 0 expectancy and a predetermined value for the standard deviation,

$$x_i^{Mut} = x_i + r_\Phi .$$

11. The method according to the preceding claim, wherein probability density distribution (Φ) is a Gaussian distribution.

12. The method according to claim 1, wherein said stop criterion (STC) is defined by a predetermined number of repetitions of said steps b) to j).

13. The method according to claim 1, wherein said stop criterion (STC) is defined by a predetermined limit of change in fitness.

14. The method according to claim 1, comprising a step of visualizing the outputted molecules.

15. The method according to claim 1, wherein said set of molecules is held in a computerized database.

16. The method according to claim 1, comprising a step of visualizing the resulting 3-D surfaces.

17. The method according to claim 1, wherein said selected candidate molecules are suitable for chemical synthesis.

18. The method according to claim 1, whereby the molecular properties represented by said descriptors are at least two of:

- molecular weight,
- number of rotatable bonds,
- number of hydrophobic groups,
- number of hydrophilic groups,
- number of acid groups,
- number of basic groups,
- number of neutral groups,
- number of zwitter groups,
- number of heavy atoms,
- number of H-bond donors,
- number of H-bond acceptors,
- number of 1-2 dipoles,
- number of 1-3 dipoles,
- number of 1-4 dipoles.

19. The method according to claim 1, whereby the molecular properties represented by said descriptors are:

- molecular weight,
- number of rotatable bonds,
- number of hydrophobic groups,
- number of heavy atoms,
- number of H-bond donors,
- number of H-bond acceptors.

20. The method according to claim 1, whereby the molecular properties represented by said descriptors are at least two of:

- molecular weight,
- number of rotatable bonds,
- number of hydrophobic groups,
- number of heavy atoms,
- number of H-bond donors,
- number of H-bond acceptors.

21. A computer system comprising means for performing the method according to claim 1.

22. The computer system according to the preceding claim comprising means for communicating with a database comprising said set of molecules.

23. A data storage means storing a program for performing the method according to claim 1.

24. A data storage means storing a database comprising the set of molecules for use with the method according to claim 1.

25. A program for storing a database comprising the set of molecules for use with the method according to claim 1.

26. A database to be used with the method according to claim 1.

27. Method of producing molecules determined by the method according to claim 1.

28. Method according to claim 27, further comprising a final step of testing said found candidate molecules in a suitable biological assay.